Emotions Recognition Using Logistic Regression and Artificial Neural Network

Harmony Mncube

Abstract—**In the past years, significant effort has occurred in developing methods of facial expression analysis. Because most investigators have used relatively limited data sets, the generalizability of these various methods remains unknown. I describe the problem space for facial expression analysis, which includes level of description, transitions among expression, eliciting conditions, reliability and validity of training and test data, individual differences in subjects, head orientation and scene complexity, image characteristics, and relation to non-verbal behavior. We then present the local hand made dataset from various Face Expression Image stored locally, which currently includes 1000 digitized image sequences from various adult subjects of varying ethnicity, performing multiple various emotion expressions. This database is the most comprehensive test-bed for our comparative studies of facial expression analysis**

Index Terms—**Facial Expressions, Kinear regression, Feed Foward Neural Networks, python3, OpenCV**

I. Introduction

Machine learning is a computer study discipline that aims at instilling human-like intelligence into computers by training them on abilities such as computer vision, natural language processing, pattern recognition etc. This paper is based on training a computer to recognize not just your face but the emotion expressed on it. Imagine walking into your home after a long day of work, and your computer immediately knows what kind of therapeutic music to play based on how you are feeling, or when you are driving the car onboard computer is able to assess your ability to drive based on your emotions. Lets not get too comfortable as this report is limited to facial expression recognition.

Most previous work on sentiment and emotion analysis has only focused on single-label classification. Hence, in this report, we focus on the multi-label emotion classification task, which aims to develop an automatic system to determine the existence in a live video of none, one, or more out of eight emotions: the eight Plutchik [1] categories (joy, sadness, anger, fear, trust, disgust, surprise, and contempt), One of the most common approaches to addressing the problem of multi-label classification is the problem transformation. With this approach, a multi-label problem is transformed into one or more single-label (i.e., binary or multi-class) problems. Specifically, single-label classifiers are learned and employed; after that, the classifiers predictions are transformed into multi-label predictions.

II. Background and Related Work

A. Introduction

Emotions are important and meaningful aspects of human behaviour. Analyzing facial expressions and recognizing their emotional state is a challenging task with wide ranging applications. In this paper, we present an emotion recognition system, which recognizes basic emotional states in facial expressions.

B. Background

*1) **Using Support Vector Machine and Multilayer Perceptron Neural Network**:* Initially, [2] detected human faces in images using the Viola-Jones algorithm. His algorithm located and measured characteristics of specific regions of the facial expression such as eyes, eyebrows and mouth, and extracts proper geometrical characteristics form each region. These extracted features represented the facial expression and based on them a classification schema, which consists of a Support Vector Machine (SVM) and a Multilayer Perceptron Neural Network (MLPNN) was used to recognize each expressions emotional content. The classification schema initially recognizes whether the expression is emotional and then recognizes the specific emotions conveyed. The evaluation conducted on JAFFE and Kohn Kanade databases, revealed very encouraging results of average recognition accuracy of 81.6%.

*2) **Using Support Vector Machine and Active Appearance Models**:* Another approach that was proposes was the analysis of the use of Active Appearance Models (AAMs) and Support Vector Machine (SVM) classifiers in the recognition of human facial emotion and emotion intensity levels. AAMs are known as a tool for statistical modeling of object shape/appearance or for precise object feature detection. The author [3] examined their properties as a technique for feature extraction. He then analyzed the influence of various facial feature data types (shape/texture/combined AAM parameter vectors) and the size of facial images on the final classification accuracy. Then, approaches to proper C-SVM classifiers (RBF kernel) training parameter adjustment as described in his paper. Moreover, he also came up with alternative way of classification for better accuracy evaluation using the human visual system as a reference point. Unlike the usual to the approach evaluation of recognition algorithms (based on comparison of final classification accuracies), his proposed an evaluation schema which was independent to the testing set parameters, such as number, age and gender of subjects or the intensity of their emotions. He then concluded to show that his automatic system gives emotion categories for images more consistent labels than human subjects, while humans are more consistent in identifying emotion intensity level compared to our system.

*3) **Using Artificial Neural Networks:*** Speech and emotion recognition improve the quality of human computer interaction and allow more easy to use interfaces for every level of user in software applications. [2] developed the emotion recognition neural network (ERNN) to classify the voice signals for emotion recognition. His ERNN had 128 input nodes, 20 hidden neurons, and three summing output nodes and used a set of about 97932 training sets to train the ERNN with a new set of 24483 testing sets to utilize the ERNN performance. His samples tested for voice recognition were acquired from the movies ”Anger Management” and ”Pick of Destiny”.His ERNN achieves an average recognition performance of 100%. This high level of recognition suggests that the ERNN is a promising method for emotion recognition in computer applications.

From all the findings and suggested solutions motioned above, I decided to use logistic regression with the help of feed forward neural network to perform emotion recognition on images. This classifier will be of help for emotion detection in therapy sessions whether the patient is angry or happy etc and classroom evaluation whether students are participating bored or happy with the lecture.

III. Face Expression Image Database

A. Introduction

We have presented some related work of Emotion Recognition Using different models from different authors in the section above. We covered the models used and the results produced by those models. In this section we will describe our problem and the dataset to be used.

B. ***Research Hypothesis***

Facial Expressed Emotion can be predicted and classified using supervised machine learning techniques such as Logistic Regression and Artificial Neural Networks. The evaluation of my model will produce results of more than 90% accurate.

C. Description of database/dataset

The dataset used in this classification problem was gathered manual from people around *The University of Witwatersrand* Campus.

D. Database evaluation

The custome made dataset consists of face grayscale images. All face images have been size-normalized and centered in a fixed size image of 28 x 28 pixels. A grayscale image is nothing but a 2D array of integers between 0 and 255. In the original dataset each pixel of the image is represented by a value between 0 and 255, where 0 is black, 255 is white and anything in between is a different shade of grey. This dataset consists of 1000 instances of images converted into a single dimension and stored as an array of integers in an excel dataset file. The images were taken directly from the person of interest and pre-processed extracting the shape of the face with the emotion expressed and resized to a 28 by 28 dimensional image. The pre-processing step was to convert the images into grey scale so that the face can be detected and then save them in one file. The last part of this process was to convert the image to a single dimension and store it. Not all images that were taken were successfully processed and stored in our dataset, some of these images failed to pass the face recognition, therefore these images where excluded from our final labelled dataset.

E. Conclusion

We have highlighted the method used to gather the dataset, and explained the processed used to finalized and process the dataset. Therefore, 70% of the processed images was used as for training and the remaining 30% was used for testing.

IV. Modeling and consistency validations

A. Introduction

Above, we have explained the steps taken to process our dataset. In the following steps, we will span our problem space for face expressions, identify the transactions between expressed and recognised emotions, ground truth of each expressed emotion (anger,contempt, disgust ,fear, happy, neutral, sadness, surprise), image resolution and finally the evaluation of our model.

B. Problem space for face expression analysis

With few exceptions, the model used attempts to recognize emotional expressions as shown in Figure 1 through 8 (i.e., anger, contempt, disgust, fear, happy, neutral, sadness, surprise). This practice may follow from the work of [4], and more recently [3], [2], who proposed that emotion-specified expressions have corresponding prototypic facial expressions. Instead, we know that emotion more often is communicated by subtle changes in one or a few discrete facial features, such as tightening of the lips in anger or obliquely lowering the lip corners in sadness [4]. Change in isolated features, especially in the area of the eyebrows or eyelids or the eye size, is typical of paralinguistic displays, for instance, raising the brows signals greeting. To capture such subtlety of human emotion and paralinguistic communication, automated recognition of fine-grained changes in facial expression is needed. Viewing videotaped facial behavior in slow motion, trained observers can manually identify and label code all possible facial emotion displayed, which are referred to as action units and may occur individually or in combinations.

C. Transitions among expressions

Each emotion expressed can vary per individual under investigation. Therefore, the expression pixel calculated from the gray scale of each image can span the whole space of dimentionality. The following figures shows sample emotion expressions from the dataset.

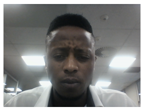


Fig. 1. anger

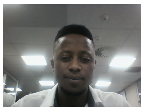


Fig. 2. contempt

Fig. 3. disgust

Fig. 4. fear

Fig. 5. happy

Fig. 6. neutral

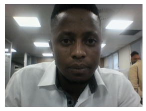


Fig. 7. sadness

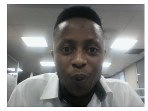
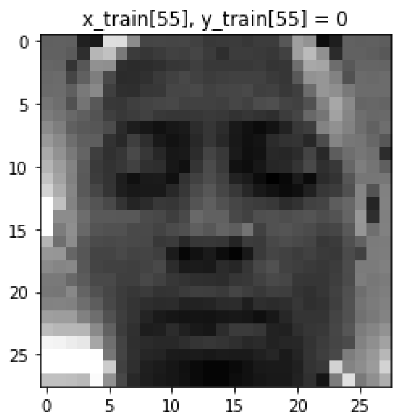


Fig. 8. surprise

D. Reliability of expression data

Some expressions where similar in contexts, therefore, the algorithm might not classify all instance correctly. These results although produced an accuracy of **TEST ACC = 0.963**.



E. Individual differences among subjects

In order for our classification to work efficiently, we model subjects of the same race, so that we don't get to have a large varience in our datasets. Therefore, we used only black people.

F. Image acquisition and resolution

After the images were processed, their resolution was reduced to 28 by 28, therefore the resulting images where pixelized. This made it easier for our model to predict outcome based on important features only. Therefore, each image can now contain only 784 features instead of the original image quality which had 49729 features.

G. Evaluation of the Model

Before I start training any of the machine learnings classification algorithm first I had to look at some constraints and performance metrics to follow :

1) Constraints:

- low-latency requirement.
- Interpretability is not that much important.
- Errors cannot be very costly.
- Probability of a data-point belonging to each class is needed.

2) Performance Metric(s):

- Multi class log-loss
- Confusion and precision matrix

Basically, the range of the log loss is [0, *inf*] and the goal of our machine learning models is to minimize this value. A perfect model would have a log loss of 0. The value of log loss increases as the predicted probability diverges from the actual label, but for a certain value of log loss other than 0, can we quantify how well is our model is performing.
Our model showed a positive performance where an error was minimised from

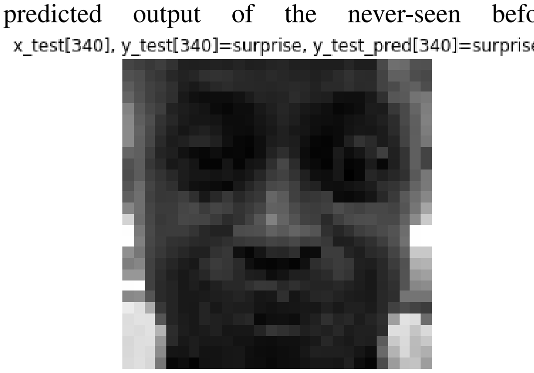
- epoch = 100, MSE = 0.414, TRAIN ACC = 0.350
- epoch = 200, MSE = 0.279, TRAIN ACC = 0.613
- epoch = 300, MSE = 0.218, TRAIN ACC = 0.741

.
.
.

- epoch = 4700, MSE = 0.020, TRAIN ACC = 0.991
- epoch = 4800, MSE = 0.019, TRAIN ACC = 0.991
- epoch = 4900, MSE = 0.019, TRAIN ACC = 0.991

From above we can see that our algorithm improved its classification minimizing the error to 0.019 with a training accuracy of 99.1%

Our test image was predicted correctly with 96% confidance. The figure below shows the predicted output of the never-seen before sample.



So far the model is doing good, Therefore, in the following I will try to find the principal component of our data. I applied the PCA to the training and then identified the number of dimentionality reduction that is possible.

Our confusion matrix looks as follows:

```
[[82  0  0  0  0  0  0  0]
 [ 4 37  0  0  1  0  0  1]
 [ 0  1 48  0  0  0  0  0]
 [ 0  0  0 37  0  0  0  0]
 [ 0  1  0  0 38  0  0  0]
 [ 0  0  0  0  0 28  0  0]
 [ 0  0  0  0  0  0 56  0]
 [ 0  2  0  0  0  0  0 53]]
```

In order to perfom PCA, we need to compute the mean vector and covariance matrix of the dataset. In previous sections, we are using row-

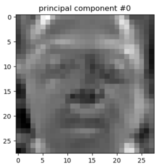


Fig. 9. pca0

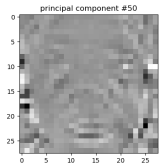


Fig. 10. pca50

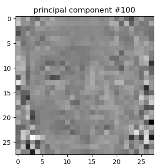


Fig. 11. pca100

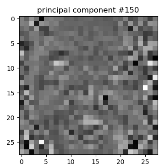
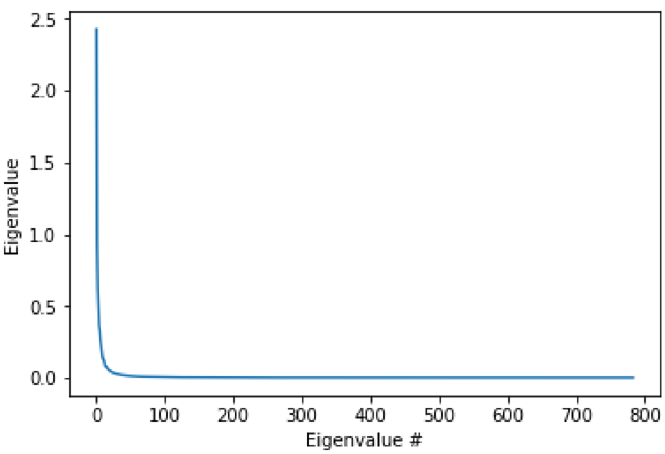
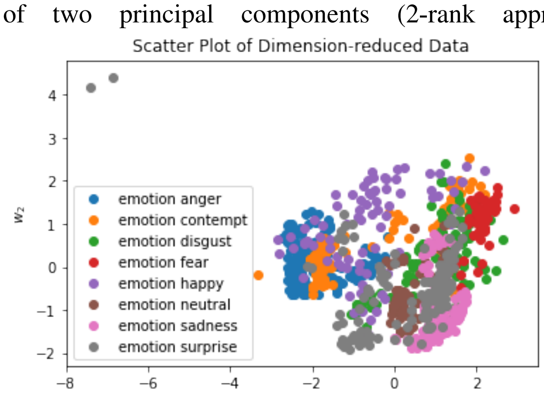


Fig. 12. pca150

major order ,therefore this is how our pca will work.



The number of principal components satisfying total variation of 0.99 is 205. As you can see above, even if we want to retain 99% of total variation, reduced dimensionality d is significantly smaller than the dimensionality $\boldsymbol{D}$ of data points.Lets display those d eigenvectors. Finally, I projected the reduced the dimensionality of a data point to d=2 by projecting it onto a vector space of two principal components (2-rank approximation).



V. Conclusion

In this work, I presented a new approach to the multi-label emotion classification task. First, I proposed a transformation method to transform the problem into a single binary classification problem. Afterwards, I developed a deep learning-based system to solve the transformed problem. The key component of our system was the embedded models namely Logistic Regression and Feed Forward Artificial Neural Networks, which used a 28 by 28 1D image to predict the emotion. Our system performed very well with an accuracy of upto 96.3%, achieving a better score that was assumed by [2] for multi-label emotion classification problem. I found that the Emotion Recognition can be improved by further finding the principal Component function can model the relationships between the input train dataset and the labels, which helps to improve the systems performance. Moreover, I showed that our system is interpretable by visualizing the principal Component spanned on a 2-d and analyzing them . These results showed that our system can perform even better since the dimension our training set can be reduced further instead of having a 28 by 28. However, some limitations have been identified. Our system does not model the relationships between faces of different sizes and the labels, because our dataset was normalized ,first cropping the image of the face and resizing it to fit our model with one contrast structure. For instance, an emotion expressed by the second person in the frame might not be determined because his face might not be close enough to be recognised as a face by the system. Thus, in our future work, I plan to work on solving this drawback. One possible solution is to adapt the attention function to model the relationships between different face in the frame and labels. SVMs may be useful in identify all objects that may look like a face and processed for validation. Moreover, I plan to work on developing a non-aggressive system that performs robustly and equally on all the emotion labels by experimenting with different ideas like using data augmentation to enrich the training data or using emotions recognised during a therapy session, party , class or even emotions expressed during sexual activities [5].

References

1 Rawls, J., "Harvard university press," in *A Theory of Justice*, March 1971, pp. 1–18.

2 Perikos, I., Ziakopoulos, E., and Hatzilygeroudis, I., "Recognize emotions from facial expressions using a svm and neural network schema," in *Engineering Applications of Neural Networks*, Iliadis, L. and Jayne, C., Eds. Cham: Springer International Publishing, 2015, pp. 265–274.

3 Beszédeš, M., Culverhouse, P., and Oravec, M., "Facial emotion classification using active appearance model and support vector machine classifier," *MG&V*, vol. 18, no. 1, pp. 21–46, Jan. 2009. [Online]. Available: http://dl.acm.org/citation.cfm?id=1643368.1643370

4 Unluturk, M. S., Oguz, K., and Atay, C., "Emotion recognition using neural networks," in *Proceedings of the 10th WSEAS International Conference on Neural Networks*, ser. NN'09. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2009, pp. 82–85. [Online]. Available: http://dl.acm.org/citation.cfm?id=1561799.1561813

5 Mirsamadi, S., Barsoum, E., and Zhang, C., "Automatic speech emotion recognition using recurrent neural networks with local attention," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 2227–2231.